

National Assessment of Educational Progress

NAEP 2012 Trends in Academic Progress

Commissioner Statement

National Center for Education Statistics

Today I am pleased to release the results of the 2012 NAEP Long-Term Trend Reading and Mathematics Assessments, the first long-term trend assessments we've conducted since 2008. These assessments were first administered in the early 1970s, giving us about 40 years of trend data.

I will start by providing an overview of NAEP. We have two basic types of assessments, what we call "Main NAEP" and "Long-term trend." Long-term trend (LTT) provides national-level results for both public and private school students, assessed by age rather than grade, at ages 9, 13, and 17. LTT provides national results in mathematics and reading for all three age levels going back to the early 1970s.

Main NAEP provides national results on student achievement in 12 subject areas. Main NAEP also provides results for all 50 states, the District of Columbia, Department of Defense schools, and selected large urban districts in mathematics and reading at grades 4 and 8, but these data do not go back nearly as far in time as long-term trend. Main NAEP results for grade 12 are more limited and, with few exceptions, are not available at the state or district level.

While the content of main NAEP has changed periodically to reflect updates in research, curriculum, assessment reform, and professional standards, the LTT assessments have remained largely unchanged. Although both main NAEP and LTT assess reading and mathematics, the assessments focus on different skills. For example, in mathematics, the topics assessed are the same—numbers, measurement, geometry, probability and statistics, and algebra. However, LTT focuses on basic skills and recall of definitions, while main NAEP goes beyond basic skills, also assessing problem solving and reasoning. Calculators are not allowed on the long-term trend assessment, but students can use them in Main NAEP for certain questions.

In reading, both assessments include a variety of text types and both require students to locate information and make inferences but main NAEP tasks tend to be more complex. The LTT reading assessments use shorter passages than the main NAEP reading assessment.

Both the long-term trend and main NAEP report scale scores but they use entirely different scales. In addition, both assessments report student achievement in terms of performance standards but different standards are used. For example, the LTT reports student performance in terms of performance levels, while main NAEP uses achievement levels (i.e. “Basic, Proficient, and Advanced”). The long-term trend performance levels, set at 50-point intervals on a 0 to 500 point scale, offer descriptions of what the students scoring at or above the levels know and can do. The long-term trend performance levels have the same meaning across all three age levels. Main NAEP achievement levels, on the other hand, are also descriptive, but are specific to each grade, and reflect judgments about what students should know and be able to do at each level. The achievement levels are set by the National Assessment Governing Board, which sets policy for NAEP.

For the long-term trend assessment, students were assessed in either reading or mathematics. In 2012 we had representative samples of close to 9,000 for each age group, for a total of more than 26,000 students per subject. We assessed 13-year-olds in the fall, 9-year-olds in the winter, and 17-year-olds in the spring. Testing time was about one hour per student.

For many years, we administered the long-term trend assessment with only limited changes, which did not affect comparability of results. In 2004 we made more significant changes, bringing the assessment up to date both in its format and in its manner of administration. In addition, we began allowing accommodations for students with disabilities and English language learners, as we do in other NAEP assessments.

In 2004, we administered both the original long-term trend and the revised assessment to different yet equivalent samples of students. Our analysis of the results determined that the changes in format and administration had no effect on student performance. Increased participation of special needs students through the use of accommodations did cause an apparent decrease in scores, but in most cases these decreases were not statistically significant. Since 2008, we have administered the revised assessment only.

The student population in the United States has changed in a variety of ways since we began measuring student performance with the long-term trend. We can use data from the long-term trend itself, going as far back as 1978, to describe these demographic changes and provide context for the changes in student performance over the years. In 1978, using 13-year-olds as an example, the student population was 80 percent White, 13 percent Black, 6 percent Hispanic, and 1 percent Asian/Pacific Islander. In 2012, the proportion of White students had decreased to 56 percent, while all of the other groups had increased—15 percent Black, 21 percent Hispanic, and 6 percent Asian/Pacific Islander.

The long-term trend lets us examine another change in student demographics—changes in the proportion of students at a given age enrolled in a given grade. For example, in 1978, 28 percent of 13-year-old students were in the seventh grade or a lower grade. In 2012, that proportion had risen to 39 percent. At the same time, the proportion attending the eighth grade fell from 72 to 60 percent. The proportion in the ninth grade or higher was 1 percent in 1978 and less than half a percent in 2012.

Why has this happened? NAEP can't tell us that, but we do know that some states have increased the age at which children may start attending kindergarten, that some parents are delaying their children's entry to school—sometimes called “redshirting”—and that some states may have changed their retention and promotion policies as well.

Reading Results

Students participating in the assessment only took a portion of the complete assessment. They read passages and responded to questions in three 15-minute sections. Each section contained three or four short passages and approximately 10 questions. The majority of the questions were presented in a multiple-choice format. Some questions and their corresponding materials were administered to more than one age group.

Scores for all three age groups are reported on a single 0–500 point scale. Because NAEP results are based on samples, there is a margin of error associated with every score. When comparing NAEP scores, we only cite differences that are larger than the margin of error—those that are statistically significant.

Nine-year-old students had an average score of 221 in 2012, higher than their score of 208 in 1971. At age 13, the score for 2012—263—was higher than the score in 1971, but it was also higher than the score for the last previous assessment, in 2008. At age 17, the score for 2012 was not significantly different from either 1971 or 2008.

The scores for groups of students ranked within an age group by level of performance can show us where improvement is taking place. We examined student performance according to the five percentiles, comparing 2012 with 1971 only. At age 9, all five groups showed increases, but the lower-performing students, at the 10th, 25th, and 50th percentiles, had larger gains—19 or 15 points, as compared to 10 or 5 points for the higher-scoring students.

Item maps and performance levels provide another perspective for interpreting long-term trend results. An item map displays the kinds of questions students at various points along the scale would be likely to be able to answer correctly and how scores compare to set performance levels. For example, students with a score of 201 were likely to be able to correctly answer a question

that required him or her to connect explicit details to recognize the main idea of an expository passage.

At age 13, as at age 9, there were long-term gains for students at all five percentiles. These gains ranged from 6 to 9 points. But there were also short-term gains, from 2008. These gains were limited to students at the 25th, 50th, and 75th percentiles. At each of the three percentiles, the increase was 3 points.

At age 17, there were long-term gains for students at the 10th and 25th percentiles—7 points at the 10th percentile and 4 points at the 25th. There was a short-term gain as well, but only for students at the 10th percentile—an increase of 5 points.

Performance levels describe a general set of knowledge and skills associated with various levels of the reading scale. For example, students scoring at Level 250 demonstrated the ability to search for specific information, interrelate ideas, and make generalizations based on what they read. In 1971, at age nine, 91 percent of students scored at level 150 or higher, while 59 percent scored at level 200 or higher, and 16 percent scored at level 250 or higher. The percentages were higher for these three performance levels in 2012 than in 1971. From 2008 to 2012, there were no increases.

At age thirteen, 93 percent of students scored at level 200 or higher, while 58 percent scored at level 250 or higher, and 10 percent scored at level 300 or higher. The percentages were higher for all three performance levels in 2012. From 2008 to 2012, there was one increase, in the percentage at or above level 250, from 58 to 66 percent.

At age seventeen, 79 percent scored at or above level 250, 39 percent scored at level 300 or higher, and 7 percent scored at level 350 or higher. When comparing 2012 to 1971, the only increase occurred in the percentage of students at or above level 250, which increased from 79 to 82 percent.

We compare the scores for 13-year-old White and Hispanic students. The first year for which we have separate results for Hispanic students is 1975. The 21-point gap between the two groups in 2012 is narrower than both the 30-point gap in 1975 and the 26-point gap in 2008. In both cases, the gap narrowed because of larger gains for Hispanic students than White students (a 17-point increase since 1975 and a 7-point increase since 2008).

Scores for White, Black, and Hispanic students all increased from the first assessment to 2012, but Black and Hispanic students had larger gains than White students at all three ages. For example, scores for White, Black, and Hispanic 9-year-old students increased by 15, 36, and 25 points, respectively from the first assessment to 2012. At ages 13 and 17, the pattern was the same—larger increases for Black and Hispanic students. White students still had higher reading scores than Black or Hispanic students.

Female students have historically had higher scores in reading than male students. The reading gender gap narrowed at age 9 only, falling from 13 points in 1971 to 5 points in 2012. Scores for both female and male students were higher in 2012, but the 17-point increase for male students was large enough to reduce the gap by 7 points.

We examined the performance of 17-year-old students over time according to the grade in which they were enrolled at age 17. In 1971, 17-year-olds who were in the 10th grade or below had an average score of 238. In 2012, the score was 266 for this group, an increase of 28 points. The proportion of 17-year-olds in or below the 10th grade rose from 14 percent in 1971 to 23 percent in 2012.

The score in 1971 for 17-year-olds in the 11th grade was 291. In 2012, it was 293, not significantly different from 1971. The proportion of 17-year-olds in the 11th grade was 73 percent in both 1971 and 2012. The average score for 17-year-olds in the 12th grade fell from 303 in 1971 to 291 in 2012. The proportion of 17-year-olds in the 12th grade fell from 13 percent in 1971 to 4 percent in 2012.

Mathematics Results

Next we will explore the results for the LTT mathematics assessment. Content included numbers, measurement, geometry, probability and statistics, and algebra. At each age level, students were assessed on their

- knowledge of basic mathematical facts and formulas,
- ability to carry out computations, and
- ability to apply mathematics in the context of daily living.

Students participating in the assessment responded to questions in three 15-minute sections. Each section contained approximately 21 to 37 questions. The majority of questions students answered were presented in a multiple-choice format. Students did not use calculators or manipulatives such as rulers during the assessment.

We discuss mathematics results from two points in time—1973 and 1978. While the first LTT mathematics assessment was administered in 1973, very few of the items were included in subsequent assessments. Thus, 1978 is the primary start of the long-term trend in mathematics. However, NCES was able to extrapolate data to compare the average scores of the 1973 assessment with later assessments so some comparisons can be made to 1973.

Scores for both 9- and 13-year-olds were higher in 2012 than in 1973. The score for 9-year-olds was 219 in 1973 and 266 in 2012; for 13-year-olds, scores increased from 266 to 285, respectively. At age 13 only, the 2012 score is also higher than the score for the previous

assessment, in 2008, and in fact for all previous assessments. At age 17, the score for 2012 was not significantly different from either 1973 or 2008.

Average mathematics scores increased for 9-year-old students at the 10th, 25th, 50th, 75th, and 90th percentiles from 1978 to 2012. The score increases were at least 22 points for students at all five percentiles. None of the percentiles showed increases in comparison with 2008.

At age 13, there were long-term gains of at least 16 points for students at all five percentiles. More specifically, the score for students at the 10th percentile in 2012 was higher than the score for students at the 25th in 1978—240 compared to 238. The increases for students at the 10th and 25th percentiles were larger than the increases at the 75th and 90th percentiles. There were also short-term gains as well. These gains were limited to students at the 75th and 90th percentiles.

As with reading, the item maps and performance levels for mathematics display the kinds of questions students at various points along the scale would likely be able to answer correctly and how scores compare to set performance levels. For example, students scoring at 240 would be likely to correctly answer a question asking them to compute the perimeter of a square. Those at 310 would likely be able to correctly rewrite an algebraic expression.

At age 17, there were long-term gains for students at the 10th, 25th, and 50th percentiles; the gains ranged from 6 to 12 points. There were no significant changes in scores from 2008.

As I have mentioned, the performance levels describe more general skills and knowledge. At level 250, students demonstrate an understanding of the four basic operations. By level 300, students are developing an understanding of number systems.

At all three ages we see increases in the percentages at the higher performance levels, comparing 2012 to 1978. At age nine, 97 percent of students scored at level 150 or higher, while 70 percent scored at level 200 or higher, and 20 percent scored at level 250 or higher. The percentages were higher for all three performance levels in 2012; the percentage at level 250 more than doubled. From 2008 to 2012, there were no increases.

At age thirteen, 95 percent of students scored at level 200 or higher, while 65 percent scored at level 250 or higher, and 18 percent scored at level 300 or higher in 1978. Again, all the percentages were higher in 2012. From 2008 to 2012, there was one increase, in the percentage at or above level 250, from 30 to 34 percent.

At age seventeen, 92 percent scored at or above level 250, 52 percent scored at level 300 or higher, and 7 percent scored at level 350 or higher. When comparing 2012 to 1978, the only increases occurred in the percentage at or above level 250, from 92 to 96 percent, and at or above level 300, from 52 to 60 percent,

As in reading, scores for White, Black, and Hispanic students increased in mathematics for all three age groups when comparing 2012 to 1973. Black students showed larger score gains than White students in mathematics at all three age groups.

In mathematics, male and female students have similar scores at both ages 9 and 13: that is, there is no gender gap. At age 17, there is a gap, favoring males. This gap fell from 8 points in 1973 to 4 points in 2012. Scores increased during that time period for female students only, narrowing the gap.

Let's summarize the results: when we compare scores in 2012 with the first assessment for both reading and mathematics, we see that scores are higher for 9- and 13-year-olds, but there were no significant changes for students at age 17. Since 2008, we have the same pattern for both subjects—an increase at age 13 only. Scores increased from the first assessment for White, Black, and Hispanic students at all three ages in both subjects. Comparing 2012 to 2008, however, we see only one increase, for Hispanic students in reading at age 13, a 7-point gain. There were no declines.

There is more information for all three age groups and both subjects in the long-term trend report card and lots more additional information available from the NAEP website. In conclusion, I would like to thank the students and schools who participated in the LTT assessments. We greatly appreciate their efforts and willingness to help inform the nation about what students know and can do in reading and mathematics.